

# Assessing Forecast Uncertainty in the National Digital Forecast Database

Matthew R. Peroutka, Gregory J. Zylstra, and John L. Wagner  
Meteorological Development Laboratory  
Office of Science and Technology  
National Weather Service, NOAA  
Silver Spring, Maryland

## 1. INTRODUCTION

NOAA's National Weather Service (NWS) has implemented a National Digital Forecast Database (NDFD) that provides its customers and partners access to gridded forecasts of sensible weather elements (e.g., cloud cover, maximum temperature). As described by Glahn and Ruth (2003), the NDFD contains a seamless mosaic of digital forecasts produced by NWS field offices working in collaboration with the National Centers for Environmental Prediction (NCEP). Table 1 lists the NDFD weather elements as well as their operational status at this time. Customers and partners use NDFD forecasts to create a wide range of text, graphic, gridded, and image products of their own.

**Table 1: NDFD Weather Elements**

<b>Operational</b>
Maximum/Minimum Temperature (MaxT/MinT)
Probability of Precipitation (PoP12)
Dew Point
Temperature
Weather
<b>Experimental</b>
Sky Cover
Quantitative Precipitation Forecast (QPF)
Wind Direction and Speed
Snow Amount
Significant Wave Height
Apparent Temperature
Relative Humidity

All NDFD weather elements except PoP12 represent single-value forecasts. The single-valued nature of the NDFD can be viewed as one of its limitations. Toth et al. (2003) observe "that all environmental forecasts are associated with uncertainty" and note that "the amount of uncertainty can be situation dependent." The NWS Strategic Plan for 2005-2010 (NWS 2005) commits the agency to "including information on forecast uncertainty to enhance customer decision processes." Consistent with this goal, the Meteorological Development Laboratory (MDL) has been investigating techniques for assessing forecast uncertainty in the NDFD and generating products from this information.

Figure 1 shows the basic structure of the NDFD UNCertainty Assessment (NUNCA). The NDFD forecast for a weather element, recent NDFD performance, and related guidance are all used to quantify expected distribution of observations for that weather element. Initial efforts have focused on MaxT and MinT. This is because these two weather elements are accessed frequently by NDFD users, and because a considerable amount of climatological data are available that describe their behavior. MDL plans to generate guidance products that allow NWS customers and partners to make better use of NDFD forecasts.

## 2. METHODS

As with other guidance techniques, NUNCA will be implemented in two distinct phases, development and implementation.

The development process begins by amassing matched pairs of forecasts (denoted by  $f$ ) and observations (denoted by  $x$ ) to form a set of developmental data. The developmental data provide input to form a model from which the joint distribution of forecasts and observations,  $p(f,x)$ , is inferred. Additional diagnostic data (denoted by  $d$ ) can be added to further refine the modeled distribution.

The implementation process uses the modeled distribution,  $p(f,x,d)$ , and current values of  $x$  and  $d$  to infer a conditional distribution of the ob-

servations given the forecast and diagnostic data,  $p(x | f, d)$ .

#### a. Data sources

The NDFD provides forecast values for regularly-spaced points on grids with mesh lengths that are close to 5 km. Efforts are underway within the NWS to routinely create a gridded "Analysis of Record" on a similar spatial scale. The NDFD and the Analysis of Record are expected to be well-matched sources of  $f$  and  $x$ . Until the analysis portion is available, however, we have been using point data as sources of  $f$  and  $x$ . Observations are taken from hourly surface reports (generally encoded in METAR); forecasts are taken from the NDFD gridpoint nearest to the verifying surface observation.

The diagnostic data that have been studied to date are Model Output Statistics (MOS; Glahn and Lowrey 1972) generated from ensemble runs of the Global Forecast System (GFS) at the NWS' National Centers for Environmental Prediction (NCEP). Erickson (1996) describes the basic processes that are used to apply MOS equations to the individual members of an ensemble run. While the Prototype MOS Ensemble Message presented in Erickson (1996) has yet to be realized, archive files that contain the MaxT, MinT, and PoP12 forecasts generated from each ensemble member are available. These archive files include 11 separate forecasts for each ENSMOS weather elements. One of the forecasts is generated by applying the MOS equations to the so-called control (unperturbed) run of the model. Five of the ENSMOS forecasts come the ensemble members that were perturbed in a "positive" way, and five of the ENSMOS forecasts come from the ensemble members that were perturbed in a "negative" way.

#### b. Transformation to percentiles

It has proved useful to transform both the forecasts and the observations from their native values to climatological percentiles. This addresses a perennial problem in modeling  $p(f, x)$ , i. e., the lack of cases in the developmental data with extreme values of  $f$  or  $x$ . One can expect this problem to be exacerbated by the short length of the NDFD's archive of forecasts (little more than one year). Another problem with modeling  $p(f, x)$  for the NDFD is the disparity of techniques used to create the NDFD grids. Human forecasters generate the NDFD using a variety of inputs, including Numerical Weather Prediction models, statisti-

cally-based forecast guidance, and observations. This can lead to significant variations in the nature of  $p(f, x)$  from region to region as well as forecast to forecast. By transforming  $f$  and  $x$  to percentiles, we hope to encourage the combining of data from multiple sites and multiple dates as  $p(f, x)$  is modeled.

#### c. Diagnostic data

Two statistics, derived from the ENSMOS MaxT guidance proved to be interesting. The first statistic is the standard deviation (SD) of the 11 MaxT forecasts contained in the ENSMOS guidance. The second interesting statistic was named the "ensemble deviation" (ED). ED is computed by differencing each of the 10 perturbed MaxT forecasts with the control MaxT, and computing the root mean square of these differences. One might argue that the 11 ensemble members should be treated as 11, equally-likely realizations taken from a single distribution, making SD the appropriate statistic. However, the statistical techniques that yield the ENSMOS forecasts did not include any perturbed model runs. Thus, the ENSMOS forecasts derived from the control can be treated as something of a standard from which the deviations of the perturbed members are measured.

#### d. NUNCA development and implementation

Figure 2 shows the NUNCA development process. An archive of NDFD forecasts and their verifying observations are gathered and transformed from their native values to percentiles. Related statistics from the associated ENSMOS forecasts are gathered as well. Together these data form a joint distribution model,  $p(f, x, d)$ , which can be used to assess the uncertainty of future NDFD forecasts.

Figure 3 shows the NUNCA implementation process. The current NDFD forecasts are transformed from their native values to percentiles. These forecasts and the associated ENSMOS statistics are then used to infer the conditional distribution  $p(x | f, d)$  from the joint distribution model,  $p(x, f, d)$ .

### 3. RESULTS

A prototype of NUNCA process was studied, using NDFD forecasts of MaxT at various forecast projection times. A set of 153 CONUS stations was selected that had long periods of record (~50 years) available in data sets provided by the

US Historical Climatology Network (USHCN; Karl, et al. 1990). Developmental data were taken from October 2004 to April 2005. Efforts focused on developing techniques that transformed MaxT forecasts and observations into climatological percentiles, assessing the nature of the joint distribution,  $p(f,x,d)$ , and evaluating SD and ED as diagnostic data.

*a. Transformation to percentiles*

For each station and at five-day intervals throughout the year, ordered observations of MaxT, taken from USHCN were used to compute frequencies for each five-day interval. Standard probability distributions were then fit to the frequencies for each interval and a cosine series was then fit to the five-day parameter values. The fits for each station were subjectively assessed, and additional terms were added to the cosine series where this addition provided a better fit. The result is a technique that can yield percentiles for any day of the year and value of MaxT as well as perform the inverse operation.

Seven open-ended probability distributions were tested for this technique. Table 2 provides the names of the distributions as well as a few subjective comments on their suitability. The distribution that fit the data best was the Generalized Lambda Distribution (GLD; Karian and Dudewicz 2000). The GLD is a powerful probability distribution that can take on a variety of shapes. This flexibility enables the GLD to model daily distributions of MaxT and MinT with impressive results. Four parameters define the GLD. Caution must be used in fitting a GLD to arbitrary data sets since it is not well-behaved for all values of its four parameters. Öztürk and Dale (1982) describe the use of GLD to model sunshine data.

**Table 2: Probability distributions tested for modeling daily distributions of MaxT and MinT**

Distribution	Variable	Comment
Normal	MaxT	Poor fit “in the tails”
Normal	ln (MaxT)	Improved fit “in the tails”
Logistic	MaxT	Better fit than either version of Normal
Laplace	MaxT	Worst fit
Gumbel	MaxT	Skewness improved fit for some stations.
Gumbel	-(MaxT)	Skewness improved fit for some stations.
Generalized		Best overall fit

**Lambda**

Figure 4 shows sample results for this technique. The figure plots the five-day frequency data for the 5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> percentiles at station BLH as well as curves produced by the percentile transform technique. The curves are able to capture a number of subtle features that can be seen in the plotted data. This is especially evident in the 5th percentile curve between days 200 and 250x.

Figure 5 compares the fitted curves from three stations that are located in drastically different climatological regimes. For each station, nine curves are plotted, one each for the 10<sup>th</sup> through 90<sup>th</sup> percentiles. Blythe, California (BLH) is located in California’s central valley, and is subject to hot summertime temperatures. Baudette, Minnesota (BDE) is located in the northern plains of the CONUS. Fort Lauderdale, Florida (FLL) is a southern, coastal station.

The percentile curves in Figure 5 clearly model a number of important characteristics of the climatology of MaxT at each station. KBLH is hot; the 90th percentile for MaxT approaches 120 °F during the summer. During the winter and spring, MaxT shows increased variability. This variability can be seen in the increased spread in the percentile curves during those seasons. By contrast, KFLH shows very little seasonal variation. During July, the spread between the 10th and 90th percentiles is remarkably small. KBDE is, by far, the coldest station of the three. The 10th percentile for MaxT drops to minus 10 °F during January. Note the large annual variation and the increased spread between the 10th and 90th percentile lines during January.

*b. Modeling the joint distribution  $p(f,x,d)$*

Computationally modeling the joint distribution  $p(f,x,d)$  can be done in very straightforward ways since the NUNCA prototype is implemented for a small number of stations. Each MaxT forecast, its verifying observation, and associated ENSMOS metric is preserved within the application. Other techniques that are less memory-intensive will likely be needed before NUNCA can be implemented for gridded NDFD forecasts.

Scatter diagrams provide one tool for qualitatively assessing the nature of  $p(f,x,d)$ . Figure 6 compares scatter diagrams for NDFD MaxT forecasts for the Day1 (24h) and Day7 (168h) time projections. The scatter diagrams certainly show a difference in the characteristics of NDFD fore-

casts for these two time projections. In each diagram, the diagonal line that runs from the lower left to the upper right represents a perfect forecast. Data points that are coincident with or near that line verify best. Visual inspection quickly suggests that Day1 forecasts verify better than Day7 forecasts as one might expect. The points on the Day7 scatter diagram cluster around the 0.50 forecast value more than the points on the Day1 diagram. This behavior coincides well with the tendency of human forecasters and objective forecasting techniques to use climatology more for later time projections. The Day7 diagram also shows fewer extreme forecasts than the Day1 diagram.

Figure 7 uses scatter diagrams to show the effects of stratifying Day1 NDFD forecasts by the ED.

Figure 8 compares the SD and ED of ENSMOS guidance as tools for understanding the performance of NDFD forecasts. The graph on the left plots the SD of the NDFD forecast error against various values of the SD of the ENSMOS. The ENSMOS SD values are binned into intervals that are 0.1 degree F wide. The graph on the right substitutes the ED of ENSMOS for the SD. In both graphs, the data cluster tightly for lower values of ENSMOS SD/ED, and become less clustered as SD/ED increase. This is expected since there are fewer cases in the developmental data with small or large values. The clustering persists in the ED graph for higher values than it does in the SD graph, suggesting that ED will be more useful in explaining the variance in NDFD forecasts. We cannot explain the apparent negative slope to the data in the right graph.

A few efforts have been made to prototype products that take advantage of the conditional distribution  $p(x | f, d)$ . These have mostly taken the form of generating a 50% confidence interval around the NDFD MaxT forecast. Other products have been considered, including a probability density function, likely expressed as the boundary values for 10-percentile intervals and the probability the variable will fall above or below certain key values (32 degrees Fahrenheit, 100 degrees, etc.).

## 12. Conclusion

The NDFD is a resource of tremendous value. One possible use of these data is to compute uncertainty information that can augment the worth of the single-valued forecasts. The NUNCA technique provides a number of tools that can derive additional value from the NDFD.

## 13. References

- Erickson, M. C., 1996: Medium-range prediction of PoP and Max/Min in the era of ensemble model output. Preprints, 15<sup>th</sup> Conf. on Weather Anal. And Forecasting, Norfolk, VA, Amer. Meteor. Soc.
- Glahn, H. R., and D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.*, **84**, 195-201.
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203-1211.
- Karian, Z. A. and E. J. Dudewicz, 2000: *Fitting Statistical Distributions*. CRC Press, 438 pp.
- Karl, T.R., C.N. Williams, Jr., F.T. Quinlan, and T.A. Boden, 1990: United States Historical Climatology Network (HCN) serial temperature and precipitation data, Environmental Science Division, Publication No. 3404, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 389 pp.
- National Weather Service, cited 2005: Working together to save lives: National Weather Service Strategic Plan 2005-2010. [Available online at [http://www.weather.gov/sp/NWS\\_strategic\\_plan\\_01-03-05.pdf](http://www.weather.gov/sp/NWS_strategic_plan_01-03-05.pdf).]
- A. Öztürk and R. F. Dale. 1982: A Study of fitting the Generalized Lambda Distribution to solar radiation data. *J. Appl. Meteor.*, **21**, 995-1004.
- Toth, Z., O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and ensemble forecasts. Forecast Verification: A Practitioner's Guide in Atmospheric Science, I. T. Jolliffe and D. B. Stephenson, Eds., John Wiley & Sons, Ltd, 137-164.

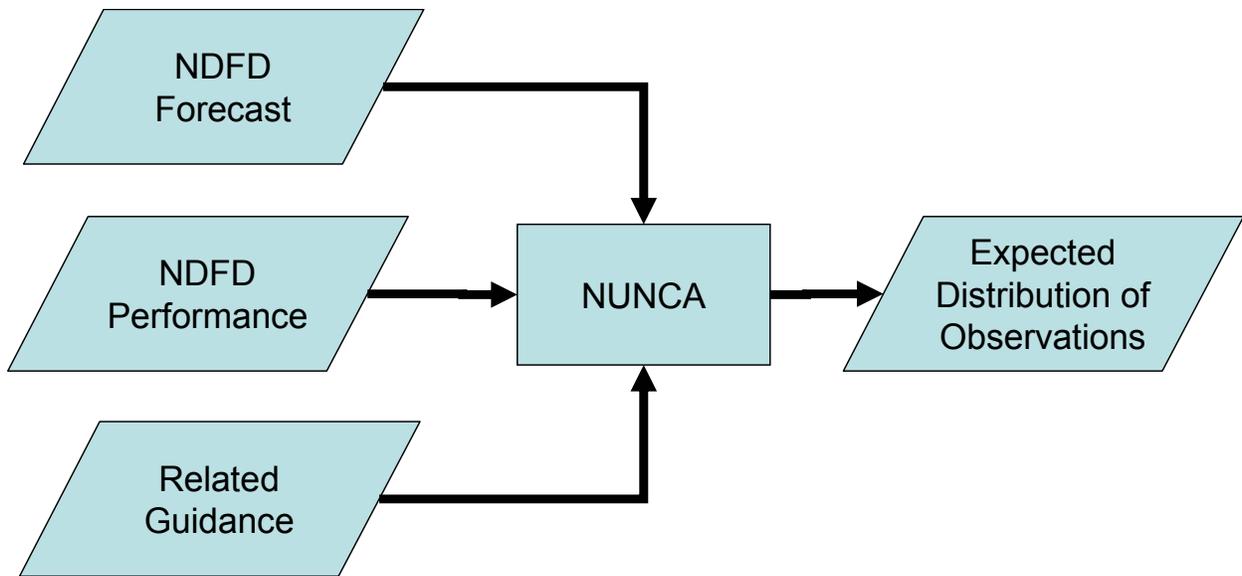


Figure 1: Overview of NUNCA process

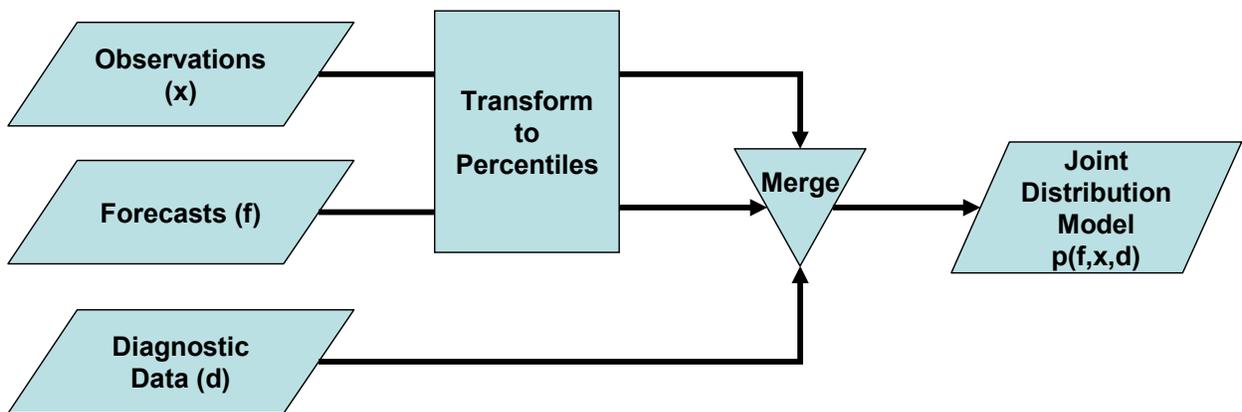


Figure 2: NUNCA development process

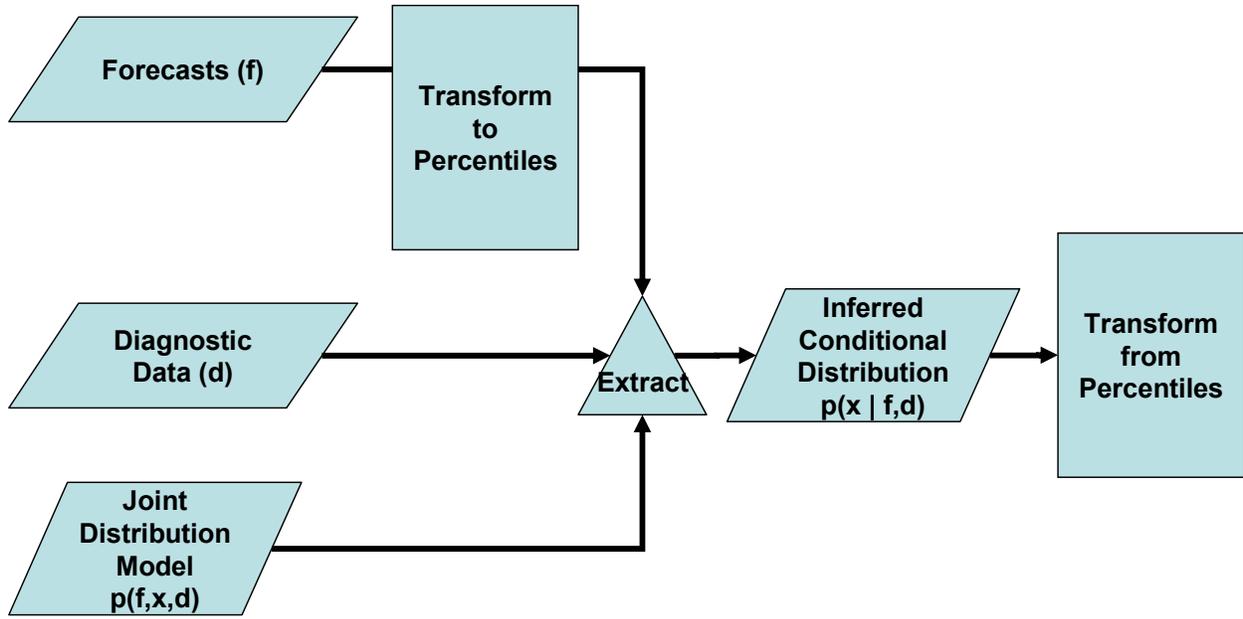
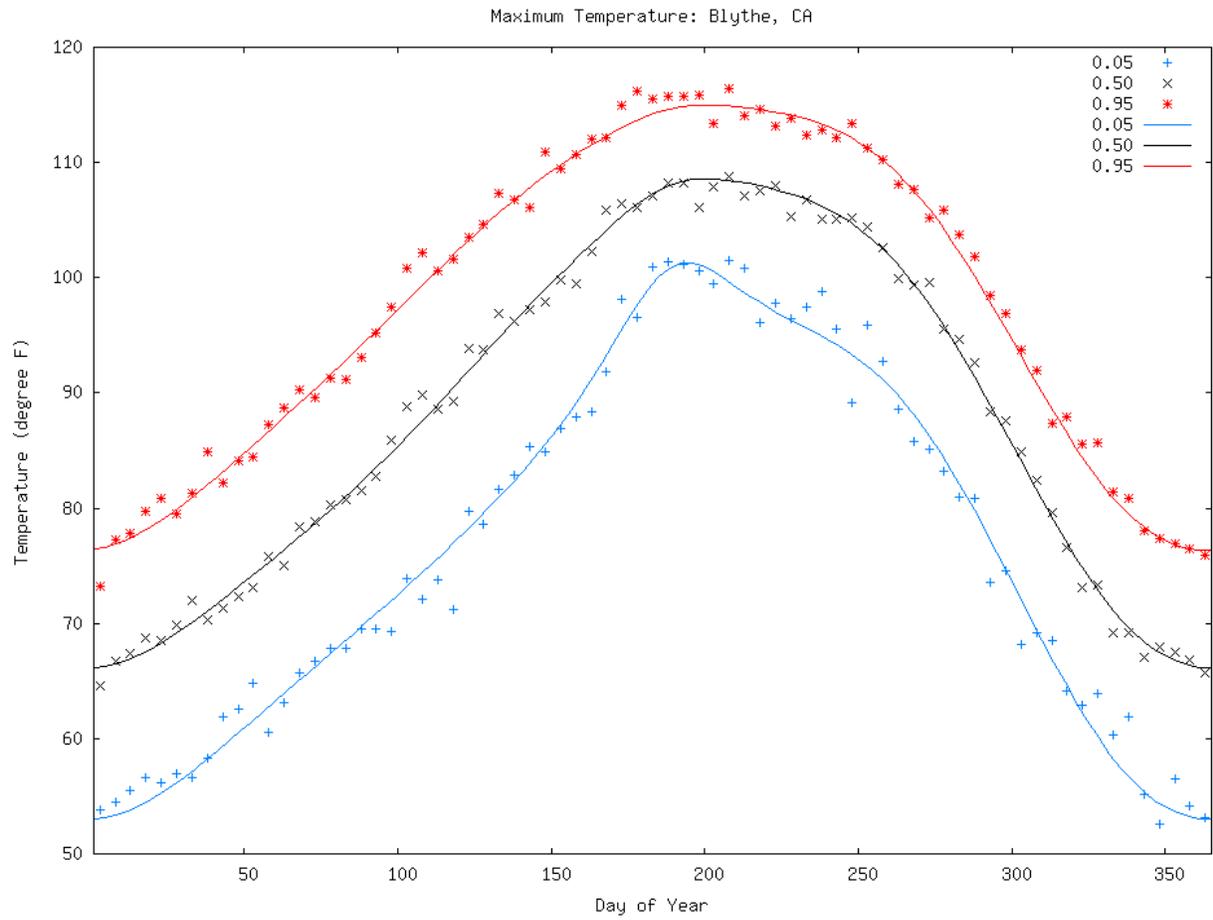
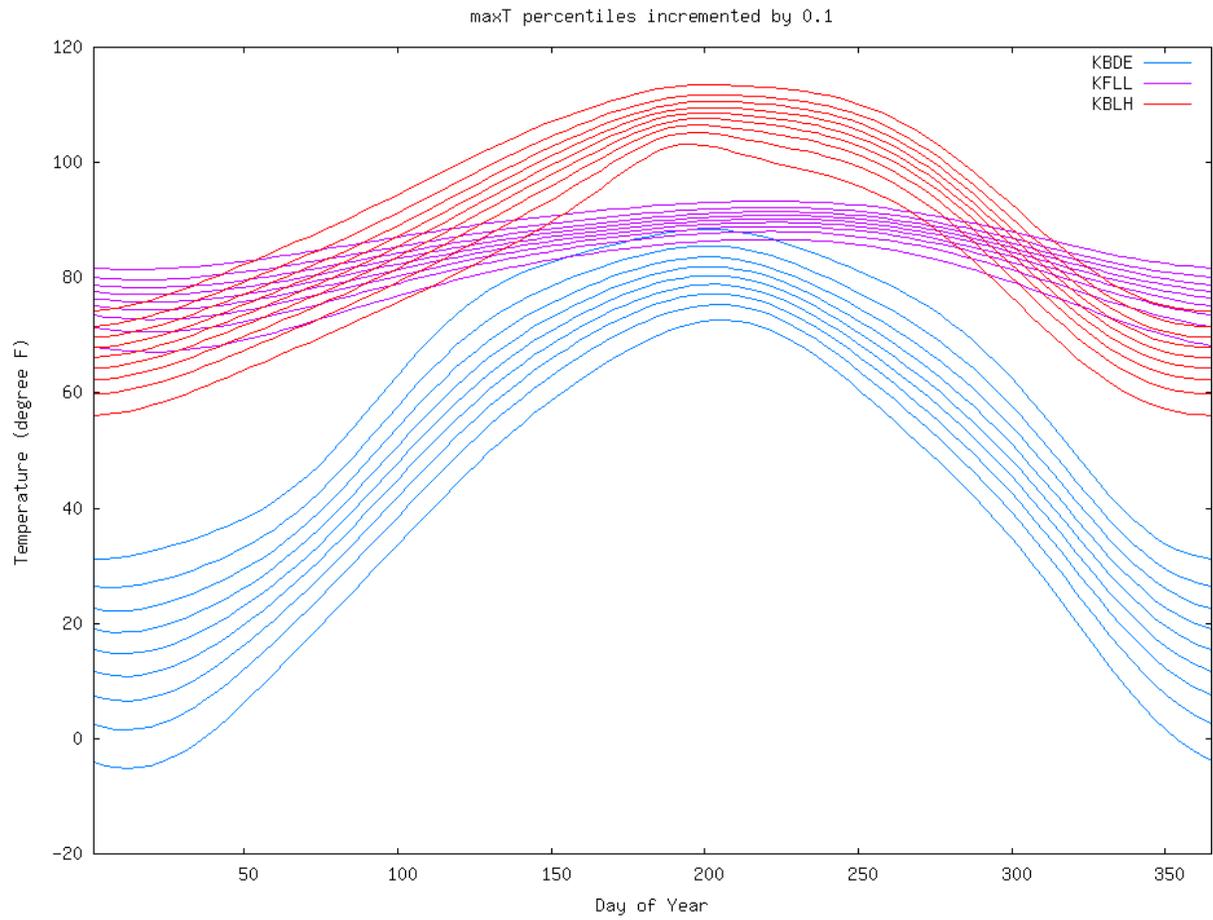


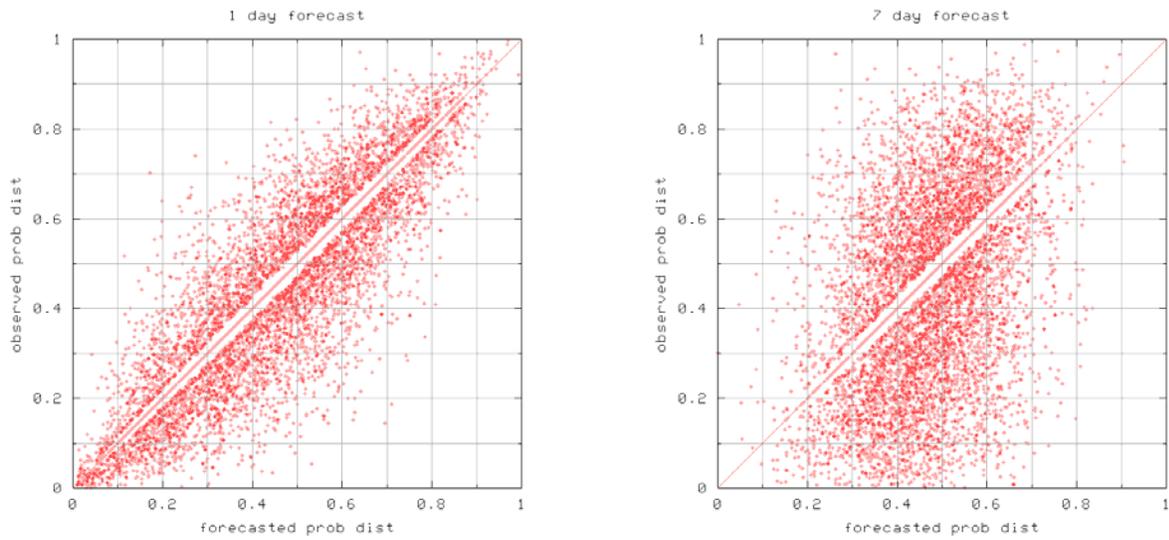
Figure 3: NUNCA implementation process



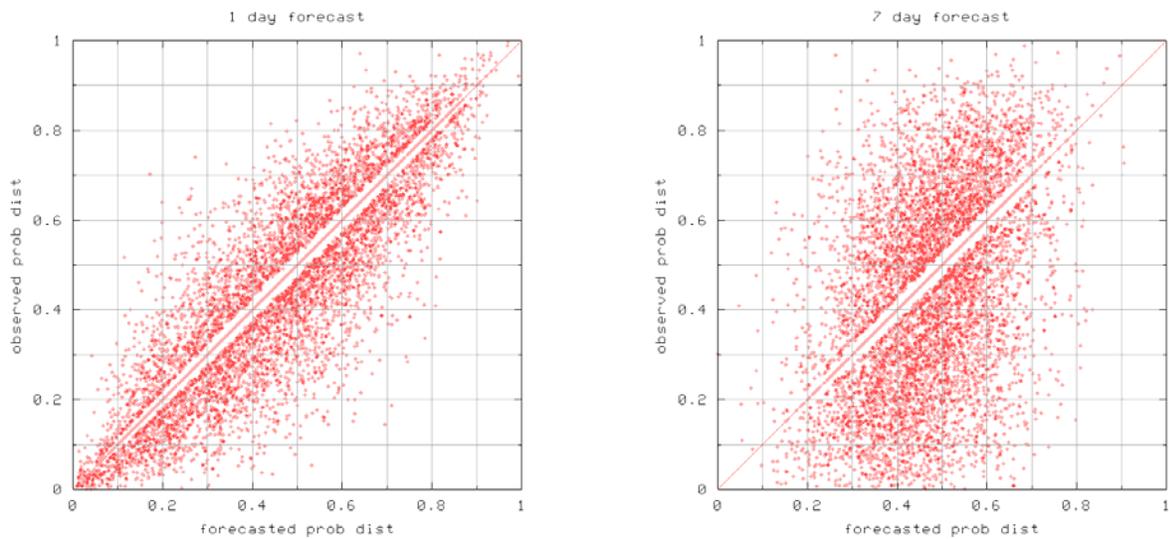
**Figure 4: Fitting five-day MaxT percentile data to a Generalized Lambda Distribution (GLD)**



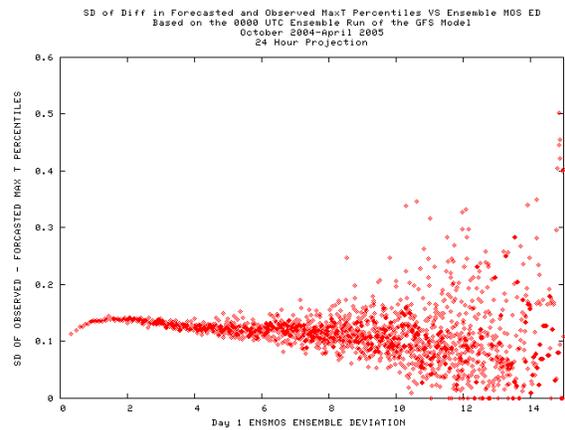
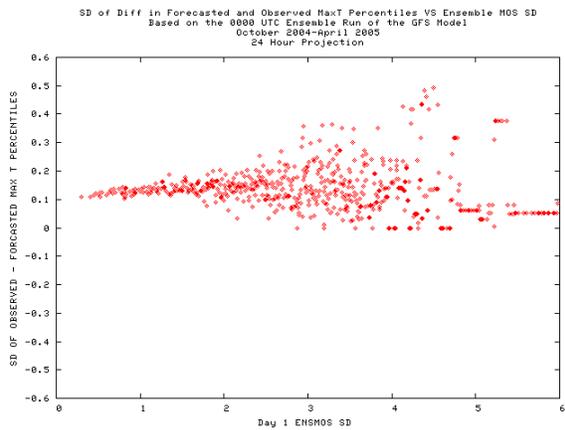
**Figure 5: Comparison of MaxT percentiles for Baudette, Minnesota (KBDE); Fort Lauderdale, Florida (KFL); and Blythe, California (KBLH)**



**Figure 6: Comparison of scatter diagrams of NDFD forecasts/verifying observations for Day1 and Day7**



**Figure 7: Comparison of scatter diagrams of NDFD forecasts/verifying observations for Day1, stratified by the ensemble deviation (defined in text) value computed from ENSMOS guidance**



**Figure 8: Graph of SD of the NDFD forecast error against various values of the SD of the ENSMOS (left) and ED of the ENSMOS (right)**

